Characterizations of Student's t-distribution via regressions of order statistics

George P. Yanev[a*] and M. Ahsanullah[b]
[a]The University of Texas - Pan American, Edinburg, Texas, USA
[b]Rider University, Lawrenceville, New Jersey, USA

October 22, 2010

Abstract

Utilizing regression properties of order statistics, we characterize a family of distributions introduced by Akhundov et al. [2], that includes the t-distribution with two degrees of freedom as one of its members. Then we extend this characterization result to t-distribution with more than two degrees of freedom.

Keywords: order statistics; characterizations, t-distribution, regression.

1 Discussion of the Results

The Student's t-distribution is widely used in statistical inferences when the population standard deviation is unknown and is substituted by its estimate from the sample. Recently Student's t-distribution was also considered in financial modeling by Ferguson and Platen [5] and as a pedagogical tool by Jones [6]. The probability density function (pdf) of the t-distribution with ν degrees of freedom (t_ν-distribution) is given for $-\infty < x < \infty$ and $\nu = 1, 2\ldots$ by

$$f_\nu(x) = c_\nu \left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2} \qquad \text{where} \quad c_\nu = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} \tag{1}$$

and $\Gamma(x)$ is the gamma function.

The vast majority of characterization results for univariate continuous distributions based on ordered random variables is concentrated to exponential and uniform families. It was not until recently, last 7-8 years, when some characterizations were obtained for t_ν-distribution with $\nu = 2$ and $\nu = 3$. In this note we communicate generalizations of these recent results for t_ν-distribution when $\nu \geq 2$. Let $X, X_1, X_2, \ldots, X_n$ for $n \geq 3$ be independent random variables with common cumulative distribution function (cdf) $F(x)$. Assume that $F(x)$ is absolute continuous with respect to the Lebesgue measure. Let $X_{1:3} \leq X_{2:3} \leq \ldots \leq X_{n:n}$ be the corresponding order statistics. Nevzorov et al. [8] (see also Nevzorov [7]) and Akhundov and Nevzorov [3] prove characterizations for the t_ν-distribution when $\nu = 2$ and $\nu = 3$,

*Corresppondng author. Email: yanevgp@ytpa.edu

respectively, assuming, in addition, $n = 3$. Here we extend these results to the general case of any $\nu \geq 2$ and any $n \geq 3$.

Let $Q(x)$ be the quantile function of a random variable with cdf $F(x)$, i.e., $F(Q(x)) = x$ for $0 < x < 1$. Akhundov et al. [2] prove that for $0 < \lambda < 1$ the relation

$$E[\lambda X_{1:3} + (1-\lambda)X_{3:3} \mid X_{2:3} = x] = x \tag{2}$$

characterizes a family of probability distributions with quantile function

$$Q_\lambda(x) = \frac{c(x-\lambda)}{\lambda(1-\lambda)(1-x)^\lambda x^{1-\lambda}} + d, \qquad 0 < x < 1, \tag{3}$$

where $0 < c < \infty$ and $-\infty < d < \infty$. Let us call this family of distributions - Q-family.

Theorem 1 (Q-family) *Assume that $E|X| < \infty$ and $n \geq 3$ is a positive integer. The random variable X belongs to the Q-family if and only if for some $2 \leq k \leq n-1$ and some $0 < \lambda < 1$*

$$\begin{aligned}\lambda E&\left[\frac{1}{k-1}\sum_{i=1}^{k-1}(X_{k:n} - X_{i:n}) \mid X_{k:n} = x\right] \\ &= (1-\lambda)E\left[\frac{1}{n-k}\sum_{j=k+1}^{n}(X_{j:n} - X_{k:n}) \mid X_{k:n} = x\right].\end{aligned} \tag{4}$$

Note that, (4) can be written as

$$\lambda E\left[\frac{1}{k-1}\sum_{j=1}^{k-1}X_{j:n} \mid X_{k:n} = x\right] + (1-\lambda)E\left[\frac{1}{n-k}\sum_{j=k+1}^{n}X_{j:n} \mid X_{k:n} = x\right] = x. \tag{5}$$

Clearly for $n = 3$ and $k = 2$, (5) reduces to (2). It is also worth mentioning here that, as Balakrishnan and Akhundov [4] report, the Q-family, for different values of λ, approximates well a number of common distributions including Tukey lambda, Cauchy, and Gumbel (for maxima).

Notice that t_2-distribution belongs to the Q-family, having quantile function (e.g., Jones [6])

$$Q_{1/2}(x) = \frac{2^{1/2}(x-1/2)}{x^{1/2}(1-x)^{1/2}}, \qquad 0 < x < 1.$$

Nevzorov et al. (2003) prove that if $E|X| < \infty$ then X follows t_2-distribution if and only if

$$E[X_{2:3} - X_{1:3}|X_{2:3} = x] = E[X_{3:3} - X_{2:3}|X_{2:3} = x]. \tag{6}$$

This also follows directly from (2) with $\lambda = 1/2$. Recall that the cdf of t_2-distribution (see Jones [6]) is

$$F_2(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{1+x^2}}\right).$$

Setting $\lambda = 1/2$ in (4), we obtain the following corollary of Theorem 1.

Corollary (t_2-distribution) *Assume that $E|X| < \infty$ and $n \geq 3$ is a positive integer. Then*

$$F(x) = F_2\left(\frac{x-\mu}{\sigma}\right) \qquad \textit{for} \quad -\infty < \mu < \infty, \quad \sigma > 0, \tag{7}$$

if and only if for some $2 \leq k \leq n-1$

$$E\left[\frac{1}{k-1}\sum_{i=1}^{k-1}(X_{k:n} - X_{i:n}) \mid X_{k:n} = x\right] = E\left[\frac{1}{n-k}\sum_{j=k+1}^{n}(X_{j:n} - X_{k:n}) \mid X_{k:n} = x\right]. \tag{8}$$

Relation (8) can be interpreted as follows. Given the value of $X_{k:n}$, the average deviation from $X_{k:n}$ to the observations less than it equals the average deviation from the observations greater than $X_{k:n}$ to it.

Remarks (i) Notice that (8) reduces to (6) when $n = 3$ and $k = 2$. (ii) Let us set $n = 2r + 1$ and $k = r + 1$ for an integer $r \geq 1$. Let $M_{2r+1} = X_{r+1:2r+1}$ be the median of the sample $X_1, X_2, \ldots, X_{2r+1}$. Then (8) implies

$$E\left[\sum_{i=1}^{r}(M_{2r+1} - X_{i:2r+1})|M_{2r+1} = x\right] = E\left[\sum_{j=r+2}^{2r+1}(X_{j:2r+1} - M_{2r+1})|M_{2r+1} = x\right].$$

If, in addition, $\bar{X}_{2r+1} = \sum_{i=1}^{2r+1} X_i/(2r+1)$ is the sample mean, then (8) reduces to Nevzorov et al. [8] t_2-distribution characterization relation

$$E\left[\bar{X}_{2r+1} \mid M_{2r+1} = x\right] = x.$$

Let us now turn to the case of t_ν-distribution with $\nu \geq 3$. Akhundov and Nevzorov [3] extend (6) to a characterization of t_3-distribution as follows. If $EX^2 < \infty$ then X follows t_3-distribution if and only if

$$E[(X_{2:3} - X_{1:3})^2|X_{2:3} = x] = E[(X_{3:3} - X_{2:3})^2|X_{2:3} = x]. \tag{9}$$

We generalize this in two directions: (i) characterizing t_ν-distribution with $\nu \geq 3$ and (ii) considering a sample of size $n \geq 3$. The following result holds.

Theorem 2 (t_ν-distribution) *Assume $EX^2 < \infty$. Let $n \geq 3$ and $\nu \geq 3$ be positive integers. Then*

$$F(x) = F_\nu\left(\frac{x-\mu}{\sigma}\right) \qquad \text{for} \quad -\infty < \mu < \infty, \quad \sigma > 0, \tag{10}$$

where $F_\nu(x)$ is the t_ν-distribution cdf if and only if for some $2 \leq k \leq n-1$

$$\begin{aligned} &E\left[\frac{1}{k-1}\sum_{i=1}^{k-1}\left(\frac{\nu-1}{2}X_{k:n} - (\nu-2)X_{i:n}\right)^2 \mid X_{k:n} = x\right] \\ &= \quad E\left[\frac{1}{n-k}\sum_{j=k+1}^{n}\left((\nu-2)X_{j:n} - \frac{\nu-1}{2}X_{k:n}\right)^2 \mid X_{k:n} = x\right]. \end{aligned} \tag{11}$$

Remarks (i) Notice that if $n = 3$, $k = 2$, and $\nu = 3$, then (11) reduces to (9). (ii) Let us set $\nu = 3$, $n = 2r + 1$ and $k = r + 1$ for an integer $r \geq 1$. If, as before, $M_{2r+1} = X_{r+1:2r+1}$ is the median of the sample $X_1, X_2, \ldots, X_{2r+1}$, then (11) implies the following equality between the sum of squares of the deviations from the sample median

$$E\left[\sum_{i=1}^{r}(M_{2r+1} - X_{i:2r+1})^2|M_{2r+1} = x\right] = E\left[\sum_{j=r+2}^{2r+1}(X_{j:2r+1} - M_{2r+1})^2|M_{2r+1} = x\right].$$

2 Proofs

To prove our results we need the following two lemmas.

Lemma 1 (Balakrishnan and Akhundov [4]) The cdf $F(x)$ of a random variable X with quintile function (3) is the only continuous cdf solution of the equation

$$[F(x)]^{2-\lambda}[1-F(x)]^{1+\lambda} = cF'(x), \qquad c>0. \tag{12}$$

Lemma 2 Let $r \geq 1$ and $n \geq 2$ be integers. Then

$$\begin{aligned} \frac{1}{k-1}\sum_{i=1}^{k-1} E\left[X_{i:n}^r \mid X_{k:n}=x\right] &= \frac{1}{F(x)}\int_{-\infty}^{x} t^r dF(t), \qquad 2\leq k\leq n; \\ \frac{1}{n-k}\sum_{j=k+1}^{n} E\left[X_{j:n}^r \mid X_{k:n}=x\right] &= \frac{1}{1-F(x)}\int_{x}^{\infty} t^r dF(t), \qquad 1\leq k\leq n-1. \end{aligned} \tag{13}$$

Proof. Using the standard formulas for the conditional density of $X_{j:n}$ given $X_{k:n}=x$ $(j<k)$ (e.g., Ahsanullah and Nevzorov [1], Theorem 1.1.1), we obtain for $r \geq 1$

$$\begin{aligned} &\frac{1}{k-1}\sum_{j=1}^{k-1} E[X_{j:n}^r \mid X_{k:n}=x] \\ &= \frac{1}{(k-1)}\frac{(k-1)}{[F(x)]^{k-1}}\sum_{j=1}^{k-1}\binom{k-2}{j-1}\int_{-\infty}^{x}[F(t)]^{j-1}[F(x)-F(t)]^{k-1-j}t^r dF(t) \\ &= \frac{1}{[F(x)]^{k-1}}\sum_{i=0}^{k-2}\binom{k-2}{i}\int_{-\infty}^{x}[F(t)]^{i}[F(x)-F(t)]^{k-2-i}t^r dF(t) \\ &= \frac{1}{F(x)}\int_{-\infty}^{x} t^r dF(t). \end{aligned}$$

This verifies (13). The second relation in the lemma's statement can be proved similarly.

2.1 Proof of Theorem 1

First, we show that equation (4) implies (3). Applying Lemma 2, for the left-hand side of (5), we obtain

$$\frac{\lambda}{k-1}\sum_{j=1}^{k-1} E[X_{j:n} \mid X_{k:n}=x]+\frac{1-\lambda}{n-k}\sum_{j=k+1}^{n} E[X_{j:n} \mid X_{k:n}=x] = \frac{\lambda}{F(x)}\int_{-\infty}^{x} tdF(t)+\frac{1-\lambda}{1-F(x)}\int_{x}^{\infty} tdF(t). \tag{14}$$

Further, since $E|X|<\infty$, we have

$$\lim_{x\to-\infty} xF(x)=0 \qquad \text{and} \qquad \lim_{x\to\infty} x(1-F(x))=0. \tag{15}$$

Therefore, integrating by parts, we obtain

$$\frac{\lambda}{F(x)}\int_{-\infty}^{x} tdF(t)+\frac{1-\lambda}{1-F(x)}\int_{x}^{\infty} tdF(t) = x-\frac{\lambda}{F(x)}\int_{-\infty}^{x} F(t)dt+\frac{1-\lambda}{1-F(x)}\int_{x}^{\infty}(1-F(t))dt. \tag{16}$$

Thus, from (14) and (16) it follows that (4) is equivalent to

$$\lambda(1-F(x))\int_{-\infty}^{x} F(t)dt = (1-\lambda)F(x)\int_{x}^{\infty}(1-F(t))dt.$$

The last equation can be written as

$$-\frac{\lambda}{1-\lambda}\int_{-\infty}^{x} F(t)dt\ \frac{d}{dx}\left[\int_{x}^{\infty}(1-F(t))dt\right] = \int_{x}^{\infty}(1-F(t))dt\ \frac{d}{dx}\left[\int_{-\infty}^{x} F(t)dt\right],$$

which leads to

$$\int_{-\infty}^{x} F(t)dt = c\left(\int_{x}^{\infty}(1-F(t))dt\right)^{-\lambda/(1-\lambda)} \qquad c>0.$$

Differentiating both sides with respect to x we obtain

$$\int_{x}^{\infty}(1-F(t))dt = c_1\left(\frac{1}{F(x)}-1\right)^{1-\lambda}, \qquad c_1>0.$$

Differentiating one more time, we have

$$[F(x)]^{2-\lambda}[(1-F(x))]^{1+\lambda} = c_2 F'(x), \qquad c_2>0, \tag{17}$$

which is (12). Referring to Lemma 1 we see that (4) implies (3).

To complete the proof of the theorem, it remains to verify that $F(x)$ with quantile function (3) satisfies (4). Differentiating (3) with respect to x we obtain

$$Q'_\lambda(x) = c(1-x)^{-(1+\lambda)}x^{-(2-\lambda)} \qquad c>0.$$

On the other hand, since $F(Q_\lambda(x)) = x$, we have $Q'_\lambda(x) = [F'(Q_\lambda(x))]^{-1}$. Therefore,

$$(1-x)^{1+\lambda}x^{2-\lambda} = cF'(Q_\lambda(x)),$$

which is equivalent to (17) and thus, to (4). This completes the proof.

2.2 Proof of Theorem 2

Notice that (11) can be written as

$$\begin{aligned}(\nu-1)x&\left[\frac{1}{n-k}\sum_{j=k+1}^{n}E[X_{j:n}\mid X_{k:n}=x]-\frac{1}{k-1}\sum_{j=1}^{k-1}E[X_{j:n}\mid X_{k:n}=x]\right]\\&=(\nu-2)\left[\frac{1}{n-k}\sum_{j=k+1}^{n}E[X^2_{j:n}\mid X_{k:n}=x]-\frac{1}{k-1}\sum_{j=1}^{k-1}E[X^2_{j:n}\mid X_{k:n}=x]\right].\end{aligned}$$

Referring to Lemma 2 with $r=1$ and $r=2$, we see that this is equivalent to

$$\begin{aligned}(\nu-1)x&\left[\frac{1}{1-F(x)}\int_{x}^{\infty}tdF(t)-\frac{1}{F(x)}\int_{-\infty}^{x}tdF(t)\right]\\&=(\nu-2)\left[\frac{1}{1-F(x)}\int_{x}^{\infty}t^2dF(t)-\frac{1}{F(x)}\int_{-\infty}^{x}t^2dF(t)\right].\end{aligned} \tag{18}$$

Let us assume that $EX=0$ and $EX^2=1$. Hence

$$\int_{x}^{\infty}tdF(t) = -\int_{-\infty}^{x}tdF(t) \quad \text{and} \quad \int_{x}^{\infty}t^2dF(t) = 1-\int_{-\infty}^{x}t^2dF(t)$$

and thus (18) is equivalent to

$$-(\nu-1)x\left(\frac{1}{1-F(x)}+\frac{1}{F(x)}\right)\int_{-\infty}^{x}tdF(t) = \frac{\nu-2}{1-F(x)}-(\nu-2)\left(\frac{1}{1-F(x)}+\frac{1}{F(x)}\right)\int_{-\infty}^{x}t^2dF(t)$$

Multiplying the above equation by $F(x)[1-F(x)]$, we find

$$-(\nu-1)x\int_{-\infty}^{x}tdF(t) = (\nu-2)\left[F(x)-\int_{-\infty}^{x}t^2dF(t)\right]. \tag{19}$$

Differentiating both sides with respect to x, we obtain

$$-(\nu-1)\int_{-\infty}^{x} t dF(t) = f(x)(x^2+\nu-2).$$

Since the left-hand side of the above equation is differentiable, we have that $f'(x)$ exists. Differentiating both sides with respect to x, we find

$$\frac{f'(x)}{f(x)} = -\frac{\nu+1}{\nu-2}\,\frac{x}{1+\frac{x^2}{\nu-2}}.$$

Integrating both sides and making use of the fact that $f(x)$ is a pdf we obtain

$$f(x) = c\left(1+\frac{x^2}{\nu-2}\right)^{-(\nu+1)/2} \quad \text{where} \quad c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi}}. \tag{20}$$

It is not difficult to see that if a random variable Z has the pdf (20), then

$$X = Z\sqrt{\frac{\nu}{\nu-2}}$$

follows t_ν-distribution, i.e., its pdf is given by (1). Thus, we have proved that (11) implies (10) when $\mu = 0$ and $\sigma^2 = 1$. The result now follows in the general case by considering the linear transformation $Y = \sigma X + \mu$.

To complete the proof, we need to verify that (11) holds when X has a cdf given by (10). If X has pdf (1) (i.e., cdf (10)), then we define

$$Z = X\sqrt{\frac{\nu-2}{\nu}},$$

which has pdf (20). Now, it is not difficult to verify that (20) satisfies (19), which in turn is equivalent to (11). The proof is complete.

References

[1] Ahsanullah, M. and Nevzorov, V.B. (2001). Ordered Random Variables, NOVA Sci. Publ., Huntington, NY.

[2] Akhundov, I.S., Balakrishnan, N., Nevzorov, V.B. (2004). New characterizations by properties of midrange and related statistics. Communications in Statistics: Theory and Methods, 33(12), 3133-3143.

[3] Akhundov, I. and Nevzorov, V.B. (2010). A simple characterization of Student's t_3 distribution. Stats. Probab. Lett., 80(5-6), 293-295.

[4] Balakrishnan, N. and Akhundov, I. (2003). A characterization by linearity of the regression function based on order statistics. Stats. Probab. Lett., 63(4), 435-440.

[5] Ferguson, K. and Platen, E. (2006). On the distributional characterization of daily log-returns of a world stock index. Appl. Math. Finance, 13(1), 19-38.

[6] Jones, M.C. (2002). Student's simplest distribution. The Statististician, 51(1), 41-49.

[7] Nevzorov, V.B. (2002). On a property of Student's distribution with two degrees of freedom. Zapiski Nauchnykh Seminarov POMI, Vol. 294, 148-157 (in Russian). (English Translation: J. Math. Sci., 2005, 127(1), 1757-1762.)

[8] Nevzorov, V.B., Balakrishnan, N., and Ahsanullah, M. (2003). Simple characterizations of Student's t_2-distribution. The Statistician, 52(3), 395-400.